Federal Aviation
Administration

DOT/FAA/AM-13/18
Office of Aerospace Medicine
Washington, DC 20591

Development, Validation, and Deployment of a Revised Air Traffic Color Vision Test: Incorporating Advanced Technologies and Oceanic Procedures and En Route Automation Modernization Systems

Thomas Chidester[1]
Nelda Milburn[1]
L. Sarah Peterson[1]
Kevin Gildea[1]
Carrie Roberts[2]
Deborah Perry[2]

[1]Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

[2]Xyant Technology, Inc.
P.O. Box 25082
Oklahoma City, OK 73125

September 2013

Final Report

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications Web site:
www.faa.gov/go/oamtechreports

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/AM-13/18		

4. Title and Subtitle	5. Report Date
Development, Validation, and Deployment of a Revised Air Traffic Control Color Vision Test: Incorporating Advanced Technologies and Oceanic Procedures and En Route Automation Modernization Systems	September 2013
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Chidester T,[1] Milburn N,[1] Peterson L,[1] Gildea K,[1] Perry D,[2] Roberts C[2]	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
[1]FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125 [2]Xyant Technology, Inc. P.O. Box 25082 Oklahoma City, OK 73125	11. Contract or Grant No.
12. Sponsoring Agency name and Address	13. Type of Report and Period Covered
Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591	14. Sponsoring Agency Code

15. Supplemental Notes

16. Abstract

Air traffic control specialists (ATCSs) are responsible for the safe, efficient, and orderly flow of traffic in the U.S. National Airspace System. Color is an integral element of the air traffic control environment. Color is used to communicate information to controllers about various modes or air traffic functions, including conflict alerts, aircraft control status, weather identification, and airspace identification.

The Aerospace Human Factors Research Division (AAM-500) of the Civil Aerospace Medical Institute developed the Air Traffic Color Vision Test (ATCOV) to determine whether individuals with color vision deficiencies (CVDs) have adequate color vision to perform critical color-related tasks involved in air traffic control. New research was required to integrate Advanced Technologies and Oceanic Procedures (ATOP, or Ocean21) and En Route Automation Modernization (ERAM) display systems into the ATCOV. The research team conducted a study to validate the addition of Ocean21 and ERAM items into ATCOV subtests.

The results of this study provided evidence of the reliability and construct validity of the revised test. It also established performance norms for subjects with normal color vision on modified subtests, determined the cut scores to apply to the revised ATCOV (version 6.1), and examined the impact of testing upon a sample of CVD subjects. Color vision ability sufficient to perform duties safely remains critical to air traffic services in the National Airspace System. Evidence of content validity for ATCS duties is provided through direct sampling of form and content of critical display data. Evidence of construct validity is provided by correlation with the Colour Assessment and Diagnosis Test threshold scores, which precisely measure color vision ability.

This resulted in a job sample test closely tied to critical task information communicated using color on air traffic displays. ATCOV 6.1 makes use of formats and color chromaticity deployed for critical information on air traffic displays, as defined by published analyses of ATCS tasks, including Ocean21 and ERAM.

17. Key Words	18. Distribution Statement
Air Traffic Control, Color Vision, Personnel Selection, Medical Qualification, Air Traffic Control Color Vision Test, En Route Automation Modernization	Document is available to the public through the Internet: www.faa.gov/go/oamtechreports

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	18	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENTS

Research reported in this paper was conducted under the Air Traffic Program Directive between the Human Factors Research Division (ANG-C1), FAA Headquarters, and the Aerospace Human Factors Research Division (AAM-500) of the FAA Civil Aerospace Medical Institute.

CONTENTS

Development, Validation, and Deployment of a Revised Air Traffic Color Vision Test:
Incorporating Advanced Technologies and Oceanic Procedures and
En Route Automation Modernization Systems

Development, Validation, and Deployment of a Revised Air Traffic Color Vision Test: Incorporating Advanced Technologies and Oceanic Procedures and En Route Automation Modernization Systems

Air traffic control specialists (ATCSs) are responsible for the safe, efficient, and orderly flow of traffic in the U.S. National Airspace System. Controllers in the cab at an airport traffic control tower (ATCT) are responsible for separating aircraft operating in close proximity to the airport and on the airport surface, including taxiways and runways. Their primary tool is direct visual surveillance of the airport area; secondary surface movement and radar displays are provided at a subset of airports. Controllers in a terminal radar approach control (TRACON) facility use radar displays to track aircraft positions in the airspace surrounding one or more airports. Controllers in air route traffic control centers (ARTCCs, or "en route centers") use radar to track and monitor positions and altitudes of aircraft flying between airports.

Color has become an integral element of the air traffic control environment. Color is used to communicate information to controllers about various modes or air traffic functions, including conflict alerts, aircraft control status, and weather. Color is used to draw attention to critical targets or urgent conditions, identify categories of information, and segment complex visual scenes (Xing & Schroeder, 2006).

The qualification standards for ATCS positions (Office of Personnel Management, undated) have long required controllers to have normal color vision (NCV). However, only rudimentary color schemes were utilized in early air traffic control (ATC) systems. The requirement of the color standard was challenged following the Americans with Disabilities Act of 1990 (though the principles of that act were incorporated for federal employment with the Rehabilitation Act of 1973). As a result, the Federal Aviation Administration (FAA) was required to develop an occupational test to determine if color vision deficient (CVD) applicants had sufficient color vision to safely accomplish job duties, despite the published standard. This allowed qualification of candidates with less than normal color vision, provided they could discriminate information critical to air traffic control that is communicated using color. Job candidate color vision is assessed in a post-offer, pre-employment medical examination. Clinical instruments, such as Pseudoisochromatic plate (PIP) tests, are used to screen applicants during the medical examination. Candidates and on-board controllers who are identified as having a CVD in a medical examination must be given an occupational test to determine if they can perform critical job duties.

In 2009, the FAA implemented a new occupational screening test (Air Traffic Color Vision Test, ATCOV 5.2; Chidester et al., 2011) for ATCS candidates tentatively selected for employment and on-board controllers who failed clinical color vision screening. On-board CVD controllers completed ATCOV testing subsequent to its deployment, and the deployment of new systems on which they must qualify will require future retesting. Following development of ATCOV, the FAA deployed two new control display systems, the Advanced Technologies and Oceanic Procedures system (ATOP; referred to hereafter as Ocean21) for oceanic airspace under United States control and the En Route Automation Modernization system (ERAM; replacing Display System Replacement/Host Computer System; DSR/HOST) for en route control centers within the continental U.S.

The research team reviewed task analysis documents specific to Ocean21 and ERAM and generalized previous task analyses to functions carried over to the new systems. Because color coding is used for critical functions, we determined that a new version of ATCOV, incorporating Ocean21 and ATCOV, would be required. Ocean21 (relative to ATCOV 5.2, covering Automated Radar Terminal System, ARTS; Standard Terminal Automation Replacement System, STARS; and DSR/HOST) adds critical colors and uses different presentation formats with additional redundant coding. ERAM uses different weather colors than DSR/HOST and adds colors for airspace and aircraft status coding. These functions were added to ATCOV in a new version hereafter called ATCOV 6.1.

Chidester et al. (2011) accomplished a preliminary linkage of ATCS tasks to color usage by using task analyses completed by Nickels, Bobko, Blair, Sands, and Tartak (1995) and updated by the American Institutes for Research (2006a, 2006b, 2006c). Chidester et al. concluded that an occupational test must ensure, for radar displays, that candidates can: (1) discriminate among datablocks coded in color to indicate whether they are under the control of the candidate (owned), are under control of someone else (unowned), are being pointed out to the candidate (pointout), or are in alert status (alert; highlighted due to potential for collision, loss of communication, hijacking, or other emergency), and discriminate these from datablocks coded for non-critical purposes (such as optional highlighting); (2) discriminate each level of weather severity communicated within a display type; and (3) detect and locate datablocks in collision alert status (conflict or low altitude) within time limitations necessary to prevent collision between an owned aircraft and another aircraft, terrain, or obstacles.

We examined the generalizability of these requirements to Ocean21 and ERAM and made reference to task analyses developed specifically for ERAM (Lockheed-Martin, 2011).

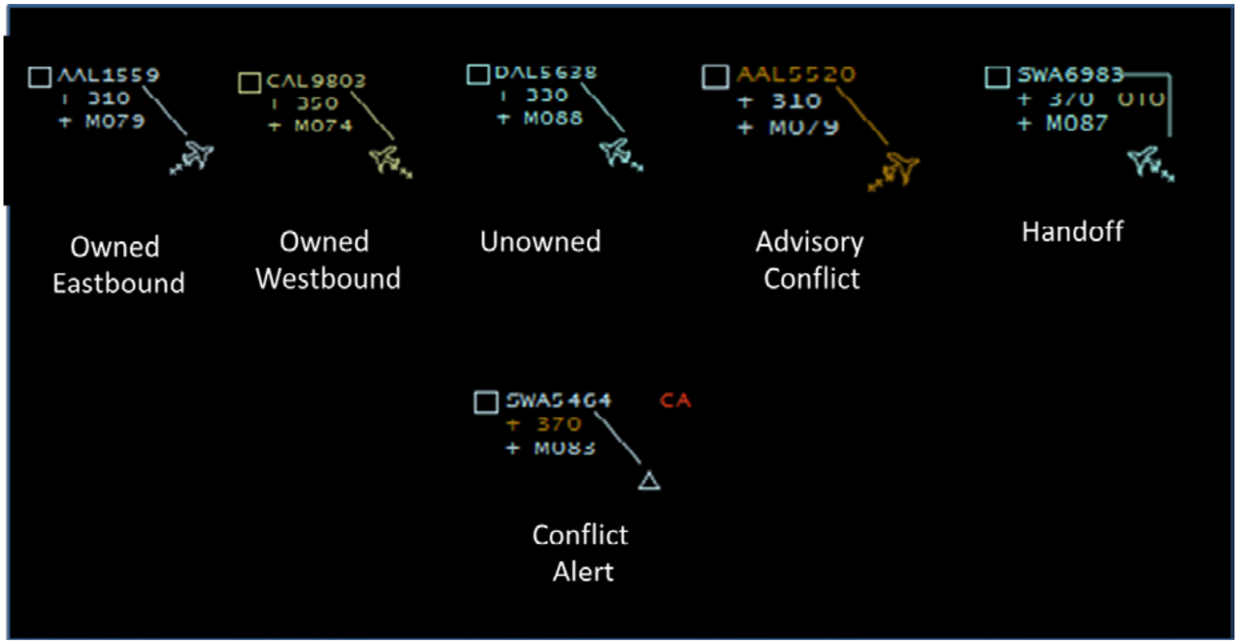

Figure 1. Ocean21. Example of Color Coding and Redundant Coding for ATCOV 6.1

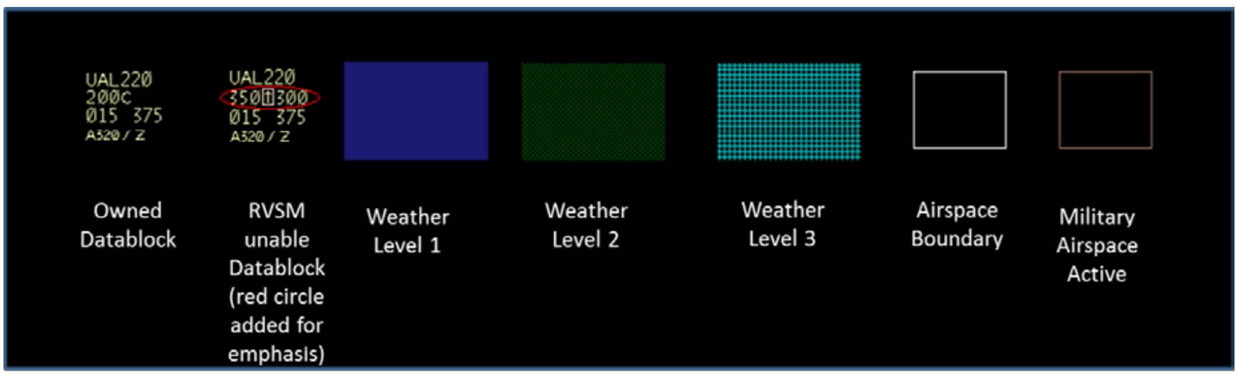

Figure 2. ERAM. Example of Color Coding for ATCOV 6.1, Including RVSM, Weather, and Military Space

Ocean21 uses color coding for owned eastbound, owned westbound, unowned, conflict alert, a)dvisory conflict, and handoff datablocks. Redundant coding is provided through an aircraft symbol indicating direction of flight and offset messages specific to handoffs and alerts ("OTO" and "CA," respectively, as represented in Figure 1).

ERAM uses color coding to indicate aircraft not capable of Reduced Vertical Separation Minima (RVSM) on their associated datablock, to indicate levels of weather, and to indicate active military airspace (Figure 2). All other uses of color were determined to be less than critical in nature by reference to the Lockheed-Martin (2011) task analysis.

This resulted in the following subtest modifications (for full functional specification, see Chidester et al., 2011):

1. Radar Identification – This subtest was modified to provide screens for Ocean21 requiring discrimination among owned eastbound, owned westbound, unowned, conflict alert, advisory conflict, and handoff datablocks and screens for ERAM requiring discrimination of RVSM unable datablocks from owned datablocks.
2. Alert Detection – This subtest was modified to add screens for Ocean21, where the subject must quickly detect a target in alert status using color and redundant coding. Redundant coding is provided by flashing red text to the right of the first line of the datablock. ERAM does not use color to indicate alert status on the radar display.
3. Weather Identification – This subtest was modified to add screens requiring discrimination among levels of weather intensity as they are color coded on ERAM radar displays. Ocean21, being a system covering offshore airspace, does not display weather radar returns.
4. Airspace Identification – This is a new subtest requiring discrimination of active military airspace boundaries from normal airspace used in ERAM.

In addition, in response to recommendations from K. Cardosi (personal communication, December 6, 2011), the background screens of the Radar Identification and Alert Detection tests were modified to require discrimination among datablocks

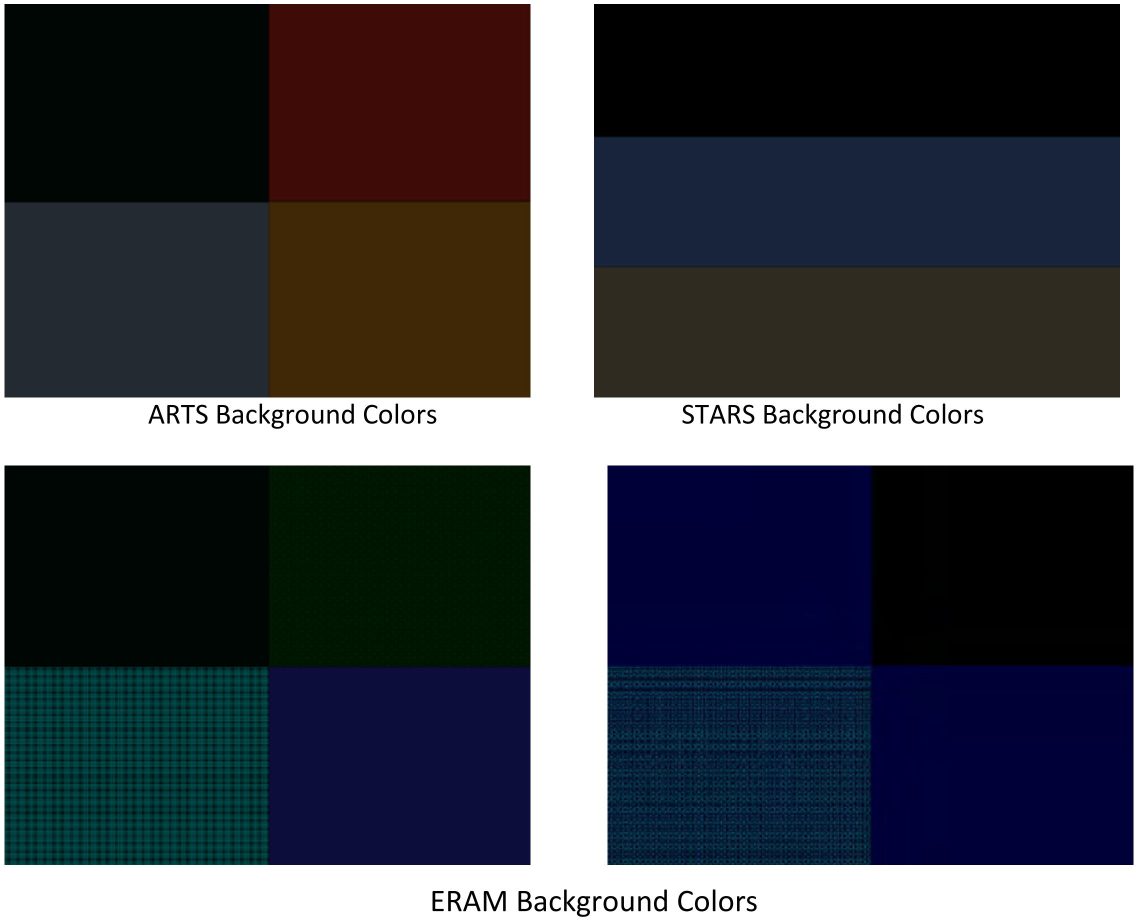

Figure 3. Examples of Color Coded Backgrounds for Each System (ARTS, STARS, ERAM) in ATCOV 6.1

superimposed upon both the native display background color and each level of weather displayed by the system. A controller may observe aircraft moving through weather on a display, but the ability to do so may be hindered by color vision deficiency. Examples of these backgrounds for each system appear in Figure 3.

For ERAM test screens, because the native background color can be adjusted from black to dark blue by the user, both colors were included in the test.

Search and distractor datablocks were randomly distributed over background colors with the proviso that search targets appear an equal number of times on each background. This ensured that subjects who pass these subtests can do so in the context of weather colors.

Luminance and chromaticity of each critical datablock, weather, or airspace color, along with background colors, were measured with a Minolta CS-200 colorimeter. Luminance (Y) and chromaticity are reported in the 1931 CID color space (xy) in Appendix A. Ocean21 displays were measured at the William J. Hughes Technical Center in Atlantic City, New Jersey, and ERAM displays were measured at the FAA Academy in Oklahoma City, Oklahoma. All displays were calibrated in accordance with field procedures. Multiple displays were measured in each location; the mean chromaticity was used as the ATCOV testing target luminance and chromaticity. Red, green, and blue color values (RGB) were then determined for each display color that would reproduce the mean field luminance and chromaticity (Yxy) value on the ATCOV display, as described in Chidester et al. (2011).

We then tested NCV and CVD subjects to assess the reliability and construct validity of ATCOV 6.1 with the addition of Ocean21 and ERAM subtest items. This also established performance norms for NCV subjects on modified subtests, determined the cut scores to apply to ATCOV 6.1, and enabled examination of the impact of testing upon a sample of CVD subjects.

METHOD

Data collection was accomplished among 98 volunteer subjects recruited from the general population of the Oklahoma City area by a subject contractor. Subjects were naive with respect to characteristics of air traffic control and were compensated for their participation. The contractor recruited 50 participants self-identified as having normal color vision, and 48 self-identified as having a color vision deficiency. Self-reports were unreliable for some subjects in each group; some NCV-identified subjects tested with mild deficiency, some CVD-identified subjects tested as normal. For analysis purposes, color vision status was measured by the Colour Assessment and Diagnosis (CAD;

Barbur, J, Rodriguez-Carmona, M., Evans, S. & Milburn, N. (2009)) test and several additional instruments listed below. By this criterion, 48 CVD (7 female) and 50 NCV (32 female) subjects participated.

Testing made use of a multi-purpose testing laboratory at the Civil Aerospace Medical Institute (CAMI). The laboratory was equipped with overhead tungsten incandescent office lights, which were illuminated during testing to produce 110 cd/m^2 at the display with a chromaticity equivalent to standard light source A. ATCOV testing was accomplished using four Dell E228WFPc 22-inch monitors calibrated to match field chromaticity using a set of RGB values unique to ATCOV. These unique values were developed using repeated measurements on two monitors, then additional monitors (including those ultimately deployed at Regional Flight Surgeon and Medical Field offices) were calibrated to match.

Prior approval for all procedures and use of human subjects was obtained from the FAA Institutional Review Board. Informed consent was obtained prior to participation, and subjects were free to withdraw from the project without consequence at any time.

All participants were between the ages of 18 and 31, the age range from which controllers may be initially hired. Subjects were screened using a Bausch and Lomb Orthorater for near and far visual acuity of 20/30 or better in both eyes, with corrective lenses, if required. All participants completed a battery of color vision tests. These tests included the signal light gun, Richmond Pseudoisochromatic Plates (4th Edition), Richmond HRR Pseudoisochromatic Plates (blue/yellow subset), Waggoner HRR Pseudoisochromatic Plates, Ishihara-38 test, Dvorine Pseudo-Isochromatic [sic] Plates, Pseudoisochromatic Plates, Ishihara Compatible (PIPIC) Color Vision test 24-Plate Edition, Cone Contrast Test, Computerized Color Vision Test (CCVT), Optec 900, and Nagel Type I anomaloscope. Additionally, several CAMI-designed experimental tests were completed, including a cockpit colors test (Pilot Color Vision Test), a paper map-reading test, map-reading using an iPAD, and identification of colored lights (incandescent and light-emitting diodes; LED). Some of these tests were administered for collateral research on pilot color vision testing and will not be discussed further in this paper.

For the present study, the primary tests of interest were the CAD (used to diagnose color vision), the previous version of ATCOV (5.2, used to compare current and former samples and performance of the present sample on previous and new versions), and ATCOV 6.1 (completed twice to assess test-retest reliability and accurately reflect field procedures). For ATCOV 5.2 and 6.1, practice opportunities were provided for each subtest and were limited by software as in the current field application. ATCOV 5.2 consists of Radar Detection, Alert Detection, and Weather Identification subtests, while ATCOV 6.1 consists of Radar Detection, Alert Detection, Weather Identification, and Airspace Identification subtests. Appendix A provides a complete set of the colors tested in ATCOV 6.1. Colors are listed by the system they represent (ARTS, STARS, ERAM, and Ocean21), the function of the color, field RGB values, color names, measured chromaticity, and a sample of the color.

For analysis purposes, color vision deficiency was determined using the CAD test. CAD diagnoses were compared to results from the Dvorine test. Subjects were classified as NCV or CVD using the CAD test, regardless of their self-identified color vision (CV) status. Subjects with red-green threshold scores on the CAD of 1.7 or less and yellow-blue threshold scores of 1.8 or less (the values at which CAD diagnoses "potential" deficiencies) were classified as NCV for the study. By these criteria, the sample included 23 deutans, one subject with a potential deutan deficiency, 18 protans, three tritans, and three subjects evidencing both red-green and yellow-blue deficiencies. Agreement between CAD classification as NCV or CVD at these cutoffs and Dvorine classification using Kappa was .86; .97 for CAD red-green (RG) diagnosis only. CAD RG threshold scores were correlated -.85 with number of correct Dvorine responses. CAD yellow-blue (YB) threshold scores were not significantly correlated with correct Dvorine responses, because Dvorine does not screen for YB deficiency. CAD thresholds and classification differ from a pass-fail classification on the Dvorine by assessing YB deficiency. In response to recommendations from K. Cardosi (personal communication, December 6, 2011), we also examined the impact on cut scores if CVD were defined as failing *any* of multiple color vision tests, rather than using only the CAD.

RESULTS

Field Specification Change, Test Programming Error, and Correction

During ATCOV 6.1 testing, air traffic facilities using Ocean21 announced a change to the color coding of unowned datablocks from gray (RGB 144,165,173; Yxy 21.93, .3411, .3515) to light sky blue (RGB 155,229,226; Yxy 28.12, .2590, .3220). Once this change was announced, our programmer created a new version of the Radar Identification and Alert Detection subtests recoding gray coding with light sky blue coding but at the same time, discovered an erroneous color reference in the Weather Identification test. The error displayed an incorrect search or distractor color on four screens, making correct versus incorrect responses to those four screens essentially random. This resulted in a significant version difference of 5.94 points of a possible 100 (d=.65, p<.01). However, there was no interaction with color vision status; the error affected both groups equally. For analysis purposes, we recalculated Weather Identification scores, removing the four affected screens for those who took the initial version. This brought with it an inability to test one STARS color on the initial version and half of all subjects but it provided the closest estimate of final version scores for affected subjects. Analysis of corrected scores showed the version difference to be reduced to 2.61 points (d=.29, p<.05), again with no significant color vision status interaction. The practical effect of the error and corrective measures was to lower the Weather Identification cut score to 80, when it could be up to 2 points higher if calculated with the reduced sample and the revised version, an error in favor of CVD subjects. The change from gray to light sky blue had no significant impact on Radar Identification and Alert Detection scores.

Characterization of Color Vision Sensitivity

The distribution of CAD RG and YB threshold scores in the sample was used to characterize the NCV and CVD groups. As expected by definition, NCV subjects had low threshold scores on both dimensions and relatively little variance (RG mean=1.26, sd=.215, min=.84, max=1.69; YB mean=1.12, sd=.248, min=.67, max=1.79). Figure 4 shows the distribution of threshold scores among NCV subjects.

CVD subjects represented a wide range of sensitivity loss in both dimensions (RG mean=15.50, sd=.8.48, min=.1.38, max=30.77; YB mean=2.12, sd=.2.94, min=.66, max=15.06). Figure 5 shows the distribution of threshold scores among CVD subjects.

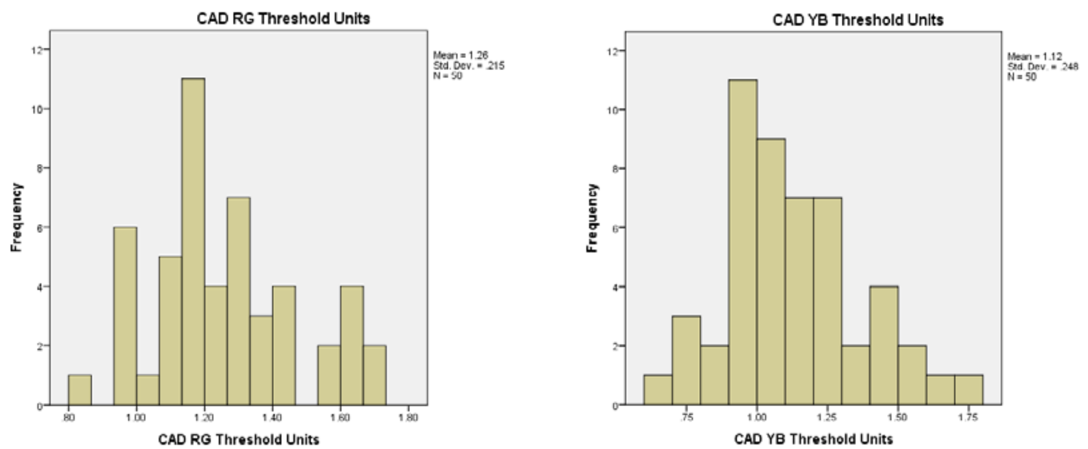

Figure 4. CAD RG and YB Threshold Scores Among NCV Subjects

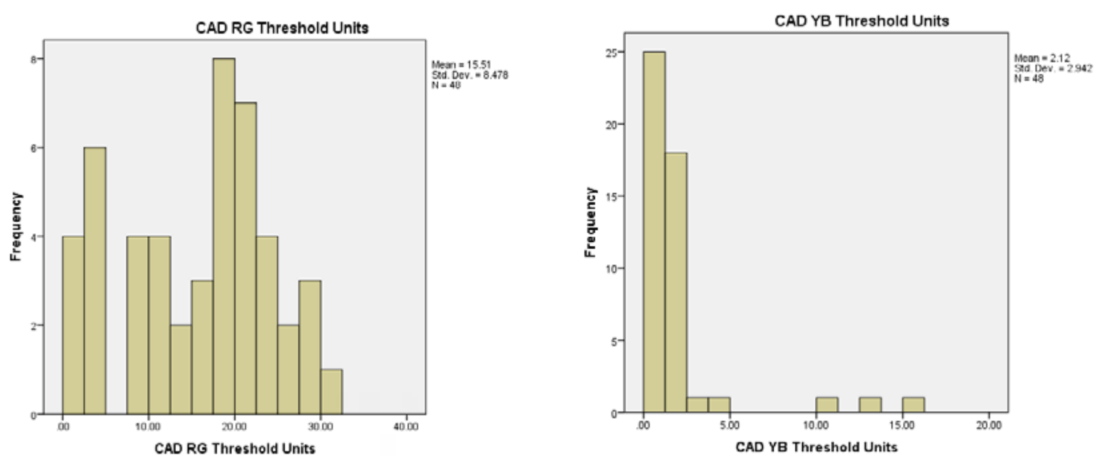

Figure 5. CAD RG and YB Threshold Scores among CVD Subjects

Test-Retest Reliability

Test-retest values were calculated for all subtests when both NCV and CVD subjects were included in a single analysis and separately for each group. The resulting values are shown in Table 1. Values across all subjects were acceptable for Radar Identification, Alert Detection, and Weather Identification. Values for Airspace Identification were low because few subjects performed poorly. Airspace Identification has questionable reliability but is necessary for content validity. For all subtests, reliability values were low for NCV subjects due to ceiling effects along a construct with which they should have no difficulty. Kappa for pass/fail on first and second attempts was .68.

Normal Color Vision Subjects

We compared air traffic naive NCV subjects who participated in this study with those tested by Chidester et al. (2011) on the version of ATCOV completed by both samples (5.2). Observed values for current subjects appear in Table 2. Fifth percentile among NCV subjects is considered because it is used as the cut score for occupational testing.

Mean scores were significantly lower in the current sample than among those in the previous study for Radar Identification (98.38 versus 99.53, t=2.78, p<.01) and Alert Detection (91.60 versus 97.76, t=6.97, p<.01), with current participants averaging .58 and 1.64 standard deviations lower, respectively. Scores for Weather Identification did not differ between the two samples. However, fifth percentile scores were lower only for Radar Identification (91 versus 95). Were this sample used for development of the cut score for the previous ATCOV version 5.2, it would have been set four points lower.

Table 1. ATCOV 6.1 Test-Retest Reliability among NCV and CVD Subjects

Subtest	*NCV*	CVD	All Participants
Radar Identification	.54	.93	.91
Alert Detection	.47	.89	.86
Weather Identification	.51	.86	.81
Airspace Identification	.00	.32	.32

Table 2. ATCOV 5.2 Descriptive Statistics Among NCV Participants

Subtest	Mean	Median	Std. Dev.	Fifth Percentile
Radar Identification	98.38	99.23	2.68	91
Alert Detection	91.60	92.00	0.76	90
Weather Identification	94.71	97.50	6.63	80

Distributions on ATCOV 6.1 among NCV subjects met expectations for each subtest, with first attempt scores concentrated at the upper range and tailing off sharply towards lower scores, as shown in Figure 6.

The mean, median, standard deviation, and fifth percentile scores for each subtest are documented in Table 3.

Cut-scores were set at the fifth percentile on the first testing attempt, except for Airspace Identification, where missing a single item would result in a score of 90, which was selected as the cut score. Using those values, 90% of NCV participants passed on the first attempt and 100% after two attempts.

Exploring recommendations from K. Cardosi (personal communication, December 6, 2011), we considered whether removing 11 participants from the NCV sample who passed CAD but failed one or more of four other color vision tests (Dvorine, Ishihara-24, Waggoner HRR, and PIPIC) would impact cut scores. This manipulation was intended to ensure the NCV sample excludes anyone with a measurable deficiency from the calculated cut scores. An unintended side effect is that it also reduces sample size to the point where the fifth percentile is the second-lowest, rather than between the second- and third-lowest score. If this method were employed, cut scores would increase on Alert Detection by two points but would decrease for Radar and Weather Identification. As a result, we rejected this method and set cut scores at 90 for Radar Identification, Alert Detection, and Airspace Identification and at 80 for Weather Identification.

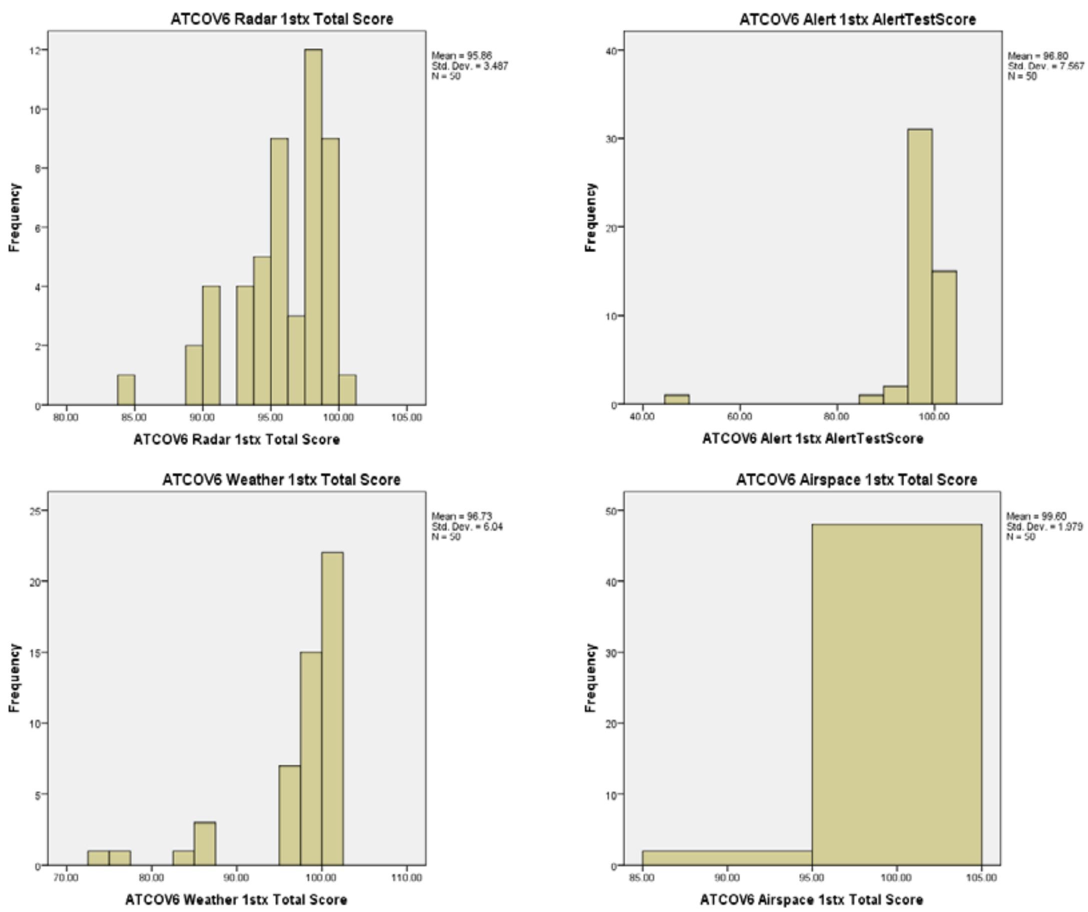

Figure 6. ATCOV 6.1 Distributions Among NCV Participants on First Attempt

Table 3. ATCOV 6.1 Descriptive Statistics Among NCV Participants

Subtest	Mean	Median	Std. Dev.	Fifth Percentile
Radar Identification	95.86	96.57	3.49	90
Alert Detection	96.80	98.00	7.57	90
Weather Identification	96.73	98.54	6.04	80
Airspace Identification	99.60	100	1.98	96

Color Vision Deficient Subjects

We compared air traffic naive CVD subjects with those tested by Chidester et al. (2011) on the version of ATCOV completed by both samples (ATCOV 5.2). Observed values for current subjects appear in Table 4.

Mean scores were significantly higher than those in the previous study for Weather Identification (92.12 versus 85.00, $t=2.59$, $p<.01$), with our participants averaging .58 standard deviations higher. No differences were found for Radar Identification or Alert Detection. Overall, 44% passed all subtests, but the test was given only once. In practice, applicants or on-board controllers taking this version are given two attempts.

We also compared CVD subjects with those tested by Chidester et al. (2011) on the CAD. Compared to the participants reported in the 2011 study, CVD participants in the current study averaged greater RG deficiency (15.51 Standard Normal Units (SNU) versus 9.54 SNU; $t=3.71$, $p<.01$) and greater variance in YB deficiency (standard deviations of 2.94 and 1.58, respectively; $F=3.47$, $p<.01$) because three current participants had deficiencies greater than 10 SNUs.

Among CVD subjects, distributions for each subtest of ATCOV 6.1 were consistent with expectations from previous research. As shown in Figure 7, on average, CVD subjects do not score as well on any subtest, but a significant proportion can discriminate critical datablocks, weatherblocks, and airspace outlines as effectively as NCV subjects (Figure 7). Comparison to Figure 6 reveals the distributions to be shifted to the left among CVD subjects. There were fewer perfect scores and substantial numbers of scores in the 60% to 80% range.

Table 4. ATCOV 5.2 Descriptive Statistics for Color Vision Deficient Participants

Subtest	Mean	Median	Std. Dev.	Pass First Attempt
Radar Identification	93.98	96.27	6.91	95
Alert Detection	84.92	87.50	8.39	44
Weather Identification	92.12	95.00	9.22	94

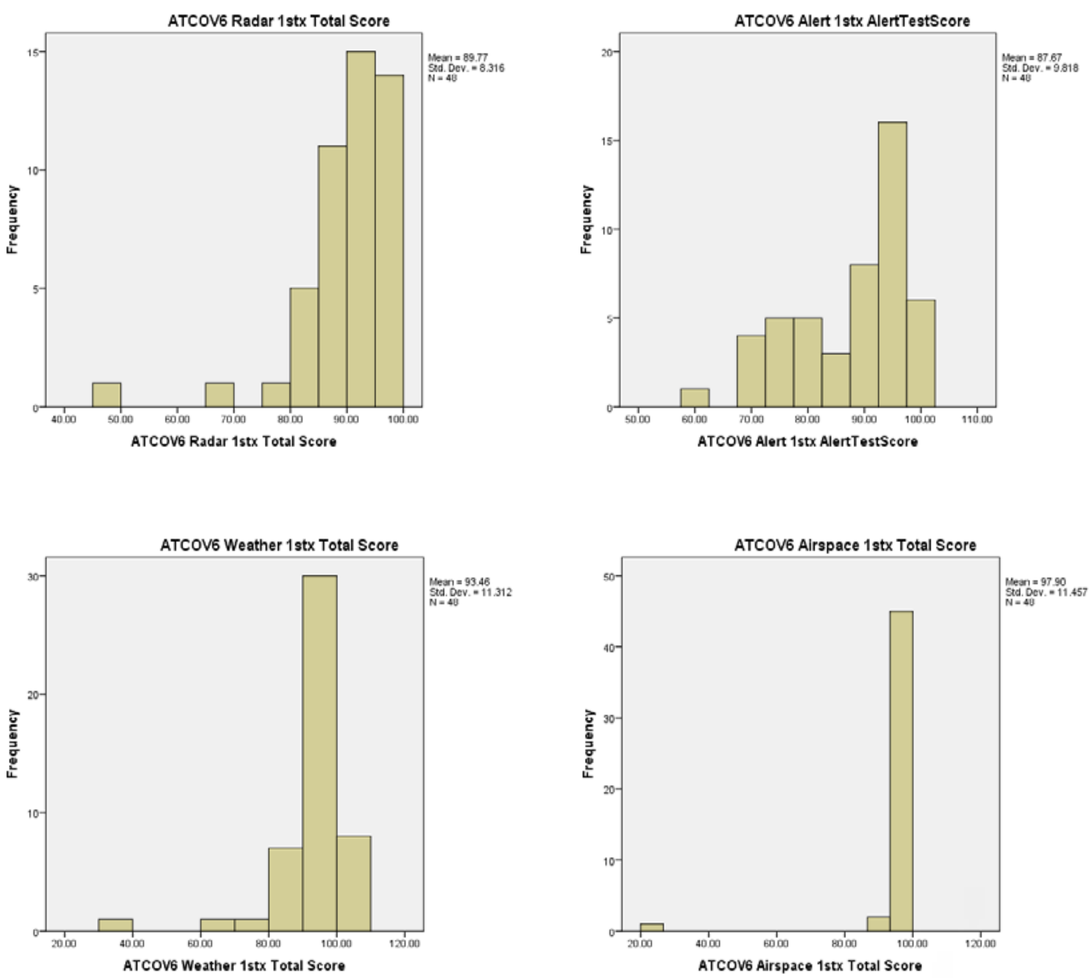

Figure 7. ATCOV 6.1 Distributions for Color Vision Deficients on the First Attempt

Airspace Identification was not as clear-cut, with only one subject receiving an extremely low score, while all others received a passing score on their first attempt.

Mean, median, standard deviation, and passing rates on the first and second attempt for each subtest on ATCOV 6.1 are documented in Table 5.

Overall, 25% of CVD subjects passed all subtests on their first attempt; 46% passed after two opportunities on all subtests. Alert Detection was the most difficult subtest. This is expected given the justifiable limitations on presentation time dictated by alert criticality. As discussed by Chidester et al. (2011), time is limited for action following an alert, so test presentation time must be similarly limited. For persons with CVD, alert coding in red, even when redundantly coded via flashing, is often not as salient, lacking a "pop-out" effect described by NCV subjects (Treisman & Gelade, 1980). The result appears to be a serial searching of targets for flashing text by CVD subjects, which cannot be completed in the test time available and, potentially, in a real emergency.

Passing rates differed by type of diagnosis: 33% of 18 protans passed after two attempts, 52% of 23 duetans passed, 100% of 3 tritans passed, and 22% of those with both RG and YB deficiencies passed.

Analyses Contrasting Subtest Performance of NCV and CVD Subjects

As shown in Table 6, for ATCOV 6.1, NCV subjects significantly outperformed CVD subjects on Radar Identification, Alert Detection, and Weather Identification, with effect sizes from .36 to .92 standard deviations on the first attempt.

Results were similar for the second attempt (Table 7). Airspace Identification does not discriminate NCV from CVD subjects due to ceiling effects but was retained in the test for content validity.

Table 5. Descriptive Statistics for ATCOV 6.1 Subtests Among CVD Participants

Subtest	Mean	Median	Std. Dev.	Pass 1st Attempt	Pass 2nd Attempt
Radar Identification	89.77	91.14	8.31	60	77
Alert Detection	87.67	91.00	9.82	54	63
Weather Identification	93.46	97.50	11.31	94	98
Airspace Identification	97.89	100	11.45	98	100

Table 6. ATCOV 6.1 Comparison of Subtest Scores by Normal Color Vision and Color Vision Deficient Participants on First Attempt

Subtest	NCV	CVD	d	p
Radar Identification	95.86	89.77	.87	<.01
Alert Detection	96.80	87.67	.92	<.01
Weather Identification	96.73	93.46	.36	<.05
Airspace Identification	99.58	97.91	.21	NS

Table 7. ATCOV 6.1 Comparison of Subtest Scores by Normal Color Vision and Color Vision Deficient Participants on Second Attempt

Subtest	NCV	CVD	d	p
Radar Identification	96.60	90.07	1.00	<.01
Alert Detection	98.16	88.93	1.17	<.01
Weather Identification	96.96	94.10	.32	<.05
Airspace Identification	99.58	97.91	.35	NS

A Kappa of .65 was obtained between CV classification by the CAD and ATCOV 6.1 pass-fail results on the first attempt, and .55 after two attempts; this appears to be associated primarily with the RG dimension (Kappa=.61 after two ATCOV attempts) and minimally with the YB dimension (Kappa=.24 after two ATCOV attempts). Chidester et al. (2011) reported a Kappa of .46 between CV normal/deficient classification by the CAD and ATCOV 5.2 pass-fail results on the first attempt, which decreased to .41 after two attempts. A Kappa value of .54 was obtained for ATCOV 6.1 with Dvorine pass/fail outcome and .08 with HRR YB pass/fail outcome, further suggesting that the primary color dimension affecting ATCOV performance is RG. Correlations between ATCOV 6.1 scores and color vision test thresholds (CAD) or number of correct responses (Dvorine, HRR YB) are shown in Table 8.

Moderately high pass/fail agreement (Kappa=.67) was observed between ATCOV 6.1 and the previous version (5.2), with disagreement observed for 13% of cases. Detailed examination of these cases revealed disagreements to primarily concern CVD subjects. While two NCV subjects failed ATCOV 5.2, both passed ATCOV 6.1. Among CVD subjects, 11 subjects passed one version but failed the other.

Analyses Examining Construct Validity

We replicated three analyses from previous studies by combining the NCV and CVD samples to assess construct validity.

We contrasted NCV versus CVD subjects using discriminant analysis of ATCOV subtest scores. The discriminant function gave greatest weight (.65) to Alert Detection, .60 to Radar Identification, .23 to Weather Identification, and .13 to Airspace Identification ($p<.01$), and correctly classified 84% of subjects by NCV versus CVD status. If subjects were scored by this function, 90% of NCV subjects and 23% of CVD subjects would be classified in a NCV group, and 10% of NCV and 77% of CVD subjects would be classified in a CVD group. This is comparable to results examining passing rates by color vision status (discussed above) and resulted in a Kappa of .67 for CV status and .61 for passing or failing ATCOV 6.1. Compared to the same analysis reported by Chidester et al. (2011) for version 5.2, 8% more subjects were correctly classified, suggesting a slightly greater sensitivity to degree of color vision deficiency for ATCOV 6.1.

We cluster-analyzed the subjects using ATCOV 6.1 subtest scores, ignoring NCV versus CVD group membership. This approach was not successful in this sample. A two-cluster solution separated one subject from all others. Adding clusters slowly separated individuals from the larger cluster. Taking a different approach, we cluster-analyzed subjects using CAD RG and YB threshold scores. A two-cluster solution grouped all NCV subjects and six CVD subjects in one group and 32 CVD subjects in a second. Agreement between cluster membership and passing or failing ATCOV 6.1 was moderately strong (Kappa=.56).

We applied factor analysis (Principal Components extraction with Varimax rotation) to the four subtest scores to assess the dimensionality of constructs underlying performance on ATCOV 6.1. Analysis of Eigenvalues and alternative numbers of extracted factors suggested that three factors accounted for response variance. A single-factor solution accounted for 47% of variance and weighted Radar Identification .94, Alert Detection .29, Weather Identification .54, and Airspace Identification .79. Given that the greater-weighted subtests were less tied to ATCOV outcomes, this approach suggested unidentified additional factors. A forced two-factor solution accounted for additional variance (73%), grouping Radar, Weather, and Airspace Identification on one factor and Alert Detection (with additional negative weighting of Weather Identification) on another. The negative weighting is not theoretically reasonable, thereby questioning the construct validity of this solution. A forced three-factor solution accounted for additional variance (95%), grouping Radar and Airspace Identification on one factor, Alert Detection on a second, and Weather Identification on a third. This solution separated color identification from color discrimination and time-limited alert detection methodologies or functions and appeared most valid.

Seeking to clarify factor structure, we added CAD red-green and yellow-blue threshold scores to a factor analysis of ATCOV subtests. A two-factor solution accounted for 64% of variance and weighted CAD RG thresholds with Alert Detection on one factor and CAD YB thresholds with Radar, Weather, and Airspace Identification on a second. This solution appeared the best fit with color vision constructs. Red-green color vision appears to be strongly tapped by time-limited detection of alerts and less so to discrimination among datablocks. Yellow-blue color

Table 8. Correlations Between ATCOV 6.1 and Color Vision Test Scores

ATCOV	CAD RG	CAD YB	Dvorine	HRR YB
Radar Identification	-.41*	-.47*	-36*	.14
Alert Detection	-.56*	-.16	-.49*	-.02
Weather Identification	-.06	-.14	.14	.29*
Airspace Identification	.12	-.45*	-.04	.08

* p < .05

vision appears associated, though less strongly, with discrimination among datablocks, weather blocks, and airspace outlines. A forced three-factor solution accounting for 80% of variance separated Weather Identification on a third factor with the other two factors essentially unchanged.

One final discriminant analysis was conducted, making use of CAD RG and YB thresholds to predict passing or failing the ATCOV 6.1 after two attempts. The discriminant function correctly classified 83% of cases, but discrimination of CVD subjects who passed from those who failed was only marginally better than chance. CAD could not fully account for ATCOV outcomes, presumably because of the way redundant coding, provided in the operational environment, was represented in the occupational test.

DISCUSSION

The purpose of this research was to add critical colors from Ocean21 and ERAM to the color palette tested by ATCOV. New research provided evidence that ATCOV 6.1 subtests were reliable, established performance norms for NCV subjects on each subtest, determined cut scores to be applied in occupational testing, and examined the impact of testing upon a sample of CVD subjects. In general, ATCOV 6.1 scores were stable among both NCV and CVD subjects, and comparable to those reported by Chidester et al. (2011) for ATCOV 5.2. Though the Airspace Identification subtest's reliability was questionable and it did not significantly discriminate NCV from CVD subjects, this was due primarily to ceiling effects, and the subtest was retained for content validity. Cut scores were set to ensure that a CVD candidate who passed the test could discriminate critical information communicated using color as well as NCV candidates. Subtest scores generally separated NCV from CVD subjects but identified fairly substantial numbers of CVD subjects who could discriminate critical information communicated using color. We concluded that ATCOV subtests adequately sampled critical information communicated using color on critical displays and function as desired and expected.

Discriminant, cluster, and factor analyses suggested that ATCOV 6.1 subtests measure color vision ability, but more than color vision is involved in successful completion. Most likely, this involves individual differences in the ability to perceive and apply redundant coding that is present in the operating environment and implemented in testing. By including precise measures of color vision ability (CAD thresholds) in our analyses, we learned that subtest scores appear tied to both red-green and yellow-blue dimensions of color vision. Subjects with protan deficiencies or a combination of RG and YB deficiencies had the greatest difficulty with the test. This provided further evidence of construct validity for the ATCOV, recognizing the mitigating impact of redundant coding.

ATCOV 6.1 was deployed to Regional Flight Surgeon and Medical Field Offices in January 2013. Chidester et al. (2011) described the need for display color standards to accompany candidate and on-board ATCS screening with occupational tests. Research to develop such a standard was initiated by the headquarters Human Factors Division in 2011 and involves active participation by Aerospace Medicine, Air Traffic, NextGen Research, and NATCA representatives. This work will determine if there exists a color set that will accommodate the largest possible cohort of CVD subjects without degrading performance or safety. As these standards come to fruition and more precise measures of field display chromaticity, luminances, and spectral distributions become available from supporting research, RGB values on ATCOV testing displays will be revised to reproduce measured field chromaticity values. K. Allendoerfer and D. Post (personal communication, May 1, 2013) completed a new round of measurements of ARTS, STARS, ERAM, and Ocean21 monitors. ARTS and STARS monitors are being refreshed with RGB LED backlit LCD monitors, which were included in this round of measurement. ATCOV RGB values were modified to conform to these updated values, (as displayed in Appendix A) in July 2013.

REFERENCES

American Institutes for Research (2006a). Air traffic control job analysis: A summary of job analytic information for air traffic en route controllers. Contractor Report. Washington, DC: Federal Aviation Administration.

American Institutes for Research (2006b). Air traffic control job analysis: A summary of job analytic information for air traffic tower controllers. Contractor Report. Washington, DC: Federal Aviation Administration.

American Institutes for Research (2006c). Air traffic control job analysis: A summary of job analytic information for air traffic TRACON controllers. Contractor Report. Washington, DC: Federal AviationAdministration.

Barbur, J., Rodriguez-Carmona, M., Evans, S., & Milburn, N. (2009). Minimum color vision requirements for professional flight crew, part III: Recommendations for new color vision standards (Report No. DOT/FAA/AM-09/11). Washington, DC: Federal Aviation Administration.

Chidester, T., Milburn N., Lomangino, N., Baxter, N., Hughes, S., & Peterson, L.S. (2011). Development, validation, and deployment of an occupational test of color vision for air traffic control specialists (Report No. DOT/FAA/AM-11/8). Washington, DC: Federal Aviation Administration.

International Commission on Illumination (1931). *Commission internationale de l'Eclairage proceedings*. Cambridge, England. Cambridge University Press.

Lockheed-Martin (2011). En route automation modernization (ERAM), Contract number DTFA01-03-C-0015, Air Traffic Training. Task and Skill Analysis Report (TASA) Job Category: Air Traffic Control. Document No. FAA-ERAM-2001-0107. Retrieved from http://www.faa.gov/air_traffic/technology/eram/

Nickels, B.J., Bobko, P., Blair, M.D., Sands, W.A., & Tartak, E.L. (1995). Separation and control hiring assessment (SACHA) final job analysis report. Contractor Report. Washington, DC: Federal Aviation Administration.

Office of Personnel Management (undated). Qualification standards, air traffic control series (2152). Washington, DC. Retrieved from: www.opm.gov/qualifications/standards/IORs/gs2100/2152. htm

Rehabilitation Act (1973). 29 U.S.C. §701. Retrieved from http://www.gpo.gov/fdsys/pkg/FR-2002-03-18/pdf/02-6536.pdf

Treisman, A.M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology,* 12, 97–136.

Xing, J. & Schroeder, D.J. (2006). Reexamination of color vision standards, Part I. Status of color use in ATC displays and demography of color-deficient controllers (Report No. DOT/FAA/AM-06/02). Washington, DC: Federal Aviation Administration.

APPENDIX A
Target Chromaticities for ATCOV as of April 2013

System	Function	Field System Specification R	G	B	Name	ATCOV Chromaticity Y	x	y	Sample
ARTS	Owned datablock	255	255	255	White	78.90	.2745	.2893	
	Unowned datablock	0	255	0	Green	60.77	.2932	.6046	
	Pointout datablock	255	255	0	Yellow	72.59	.3992	.5185	
	Alert Datablock	255	0	0	Red	15.92	.6235	.3260	
	Weather level 1	96	96	96	Dark Gray	6.35	.2615	.2673	
	Weather level 2	172	90	0	Brown	10.34	.5352	.3890	
	Weather level 3	204	48	0	Reddish Brown	9.88	.6160	.3238	
	Background	0	0	0	Black	.12	.2481	.2462	
STARS	Owned datablock	255	255	255	White	78.90	.2745	.2893	
	Unowned datablock	0	255	0	Green	60.77	.2932	.6046	
	Pointout datablock	255	255	0	Yellow	72.59	.3992	.5185	
	Alert Datablock	255	0	0	Red	15.92	.6235	.3260	
	Highlight Datablock	0	255	255	Cyan	67.66	.2105	.2880	
	Weather level 1	57	115	115	Dark Gray Blue	8.80	.2034	.2609	
	Weather level 2	124	124	64	Dark Mustard	11.45	.3582	.4622	
	Background	0	0	0	Black	.12	.2481	.2462	
ERAM	Owned datablock	238	243	174	Yellow	71.45	.3614	.3971	
	RVSM Unable DB	243	212	207	Coral	60.36	.3473	.3499	
	Weather level 1	25	25	112	Midnight Blue	2.83	.1790	.1590	
	Weather level 2	0	100	0	Green	7.56	.3165	.5550	
	Weather level 3	0	200	200	Cyan	38.52	.2320	.3475	
	Airspace boundary	100	100	100	White	6.72	.3173	.3334	
	Active airspace	144	112	96	Brown	17.50	.3859	.3675	
	Background	0	0	25	Dark Blue	.49	.2834	.2717	
Ocean21	Owned EB datablock	240	255	255	White	79.69	.3330	.3566	
	Owned WB datablock	240	230	140	Yellow	65.75	.4129	.4348	
	Unowned datablock	141	182	205	Light Sky Blue	35.79	.2635	.3206	
	Conflict Alert	255	0	0	Red	21.80	.6453	.3368	
	Advisory Conflict	255	140	0	Orange	32.90	.5705	.3905	
	Background	0	0	0	Black	.16	.3102	.2601	